

COMPONENT PART NOTICE

THIS PAPER IS A COMPONENT PART OF THE FOLLOWING COMPILATION REPORT:

TITLE: Computing Science and Statistics: Proceedings of the Symposium on Interface
Critical Applications of Scientific Computing (23rd): Biology, Engineering,
Medicine, Speech Held in Seattle, Washington on 21-24 April 1991.

AD-A252 938.

TO ORDER THE COMPLETE COMPILATION REPORT, USE _____.

- THE COMPONENT PART IS PROVIDED HERE TO ALLOW USERS ACCESS TO INDIVIDUALLY AUTHORED SECTIONS OF PROCEEDING, ANNALS, SYMPOSIA, ETC. HOWEVER, THE COMPONENT SHOULD BE CONSIDERED WITHIN THE CONTEXT OF THE OVERALL COMPILATION REPORT AND NOT AS A STAND-ALONE TECHNICAL REPORT.

THE FOLLOWING COMPONENT PART NUMBERS COMPRISE THE COMPILATION REPORT:

AD#: AD-P007 096 thru AD-P007 225
 AD#: _____ AD#: _____
 AD#: _____ AD#: _____

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC
ELECTE
JUL 23 1992
S A D

This document has been approved
 for public release and sale; its
 distribution is unlimited.



Improved Methods for Estimating Parameters in Discrete Data Analysis

DAVID J SCOTT

*School of Information and Computing Sciences
Bond University
Gold Coast, Queensland
Australia*

WANG DONG QIAN

*La Trobe University
Bundoora, Victoria
Australia*

1 Introduction

For data which take the form of a two-way contingency table, many authors have examined models other than the common model of independence of two classifications. In the literature are quasi independence models, clustered sampling models, intraclass models, and the Bradley-Terry model, among others. In order to estimate model parameters, iterative methods are required, and this leads to the problem of developing efficient algorithms.

2 Quasi Independence Models

If there are cells in a contingency table which *a priori* have a zero count, and there are cells in that row or column which have non-zero counts, the independence model is not appropriate. To cope with this problem of so-called structural zeroes, the notion of quasi independence is useful. In a quasi independence model the row and column classifications are independent, provided the cells with *a priori* zero counts are ignored.

Examples given in the literature of quasi independence models include the random pairing models of

Larntz and Weisberg (1976) and de Jong, Greig and Madan (1983), the mover-stayer model as discussed by Morgan and Titterton (1977) and the model of Lemon and Chatfield (1971), which is an alternative to a Markov chain model. Larntz and Weisberg's model can be obtained from Lemon and Chatfield's by folding along the main diagonal.

These models are all for data which take the form of a contingency table with entries on the main diagonal which are zero *a priori*, with the exception of Larntz and Weisberg's model, where the entries on or below the main diagonal are all zero *a priori*.

In order to fit these models to data, various methods have been proposed. Iterative proportional fitting (IPF) is commonly used. IPF requires that the model be log linear, which is not the case for de Jong, Greig and Madan's model. The Newton-Raphson method converges quickly, but it is not easy to implement if the Hessian is not diagonal. The Hessian is diagonal for the mover-stayer model and de Jong, Greig and Madan's random pairing model, but not for the other models. Fixed point iterations have been used by a number of authors, but these can be slow to converge. Brown (1974) developed a method for dealing with *a priori* ze-

roes, which iterates over the cells which are known to be zero. Brown's method becomes less and less efficient with an increasing number of zero cells. de Jong, Greig and Madan (1983) developed a method for fitting their random pairing model which involves a reparameterisation, then fixed point iteration. This method can also be adapted to fit other quasi independence models.

It can be shown that when the table is symmetric, the parameter estimates obtained for Lemon and Chatfield's (1971) model are identical with those obtained from fitting the mover-stayer model. This means that the models of both Lemon and Chatfield, and Larntz and Weisberg may be fitted by symmetrising the data and fitting the mover-stayer model. Thus a readily implementable Newton-Raphson approach is available for these models.

All the methods so far discussed estimate a probability distribution and involve a multivariable iteration. The authors have developed new methods for fitting quasi independence models which require the solution of a nonlinear equation in a single unknown. This equation is readily solved using Newton's method. This gives fast, very easily-implemented methods. No programming is required, and well-known packages such as Minitab or a spreadsheet may be used to do the calculations.

3 Other Models

In the clustered sampling model with clusters of size two, if a number of clusters are observed, and each member of the cluster is classified according to some characteristic, the data take the form of a two-way contingency table. Cohen (1976) has given a method which requires a two-stage iteration procedure, with one stage being the solution of a non-linear equation. Using a reparameterisation the authors were able to reduce the computations required. A two-stage iteration is still required, but only simple expressions need to be evaluated at any stage. This work is reported in Scott and Wang (1990).

Attempts were made to develop improved methods for intraclass models (see Haber(1982)) and the Bradley-Terry model, without success.

References

- [1] Cohen, J.E. (1976). The distribution of the chi-squared statistic under clustered sampling. *J. Amer. Statist. Assoc.* **71**, 665-670.
- [2] Brown, M.B. (1974). Identification of the sources of significance in two-way contingency tables. *Appl. Statist.*, **23**, 405-413.
- [3] de Jong, P., Greig, M., and Madan, D. (1983). Testing for random pairing. *J. Amer. Statist. Assoc.*, **78**, 332-336.
- [4] Haber, M. (1982). Testing for independence in intra class contingency tables. *Biometrics*, **38**, 93-103.
- [5] Larntz, K. and Weisberg, S. (1976). Multiplicative models for dyad formation. *J. Amer. Statist. Assoc.*, **71**, 455-461.
- [6] Lemon, R.E. and Chatfield, C. (1971). Organisation of song in cardinals. *Anim. Behav.* **19**, 1-17.
- [7] Morgan, B.J.T. and Titterton, D.M. (1977). A comparison of iterative methods for obtaining maximum likelihood estimates in contingency tables with missing diagonal. *Biometrika*, **64**, 265-269.
- [8] Scott, David J., and Wang Dong Qian (1990). An iterative method for estimation of parameters in a clustered sampling model. *Aust. J. Statist.*, **32**, 317-325.